What is an appropriate salary for Data Science and Machine Learning employees?

Luis Silva
February 2023 Cohort
Capstone Project for Springboard Data Science Bootcamp

The problem

With Data Science jobs having so many variables like experience level, employee location and remote/in-person modalities, it can be hard to determine the appropriate salary for an employee.

Who might care?

Employers:

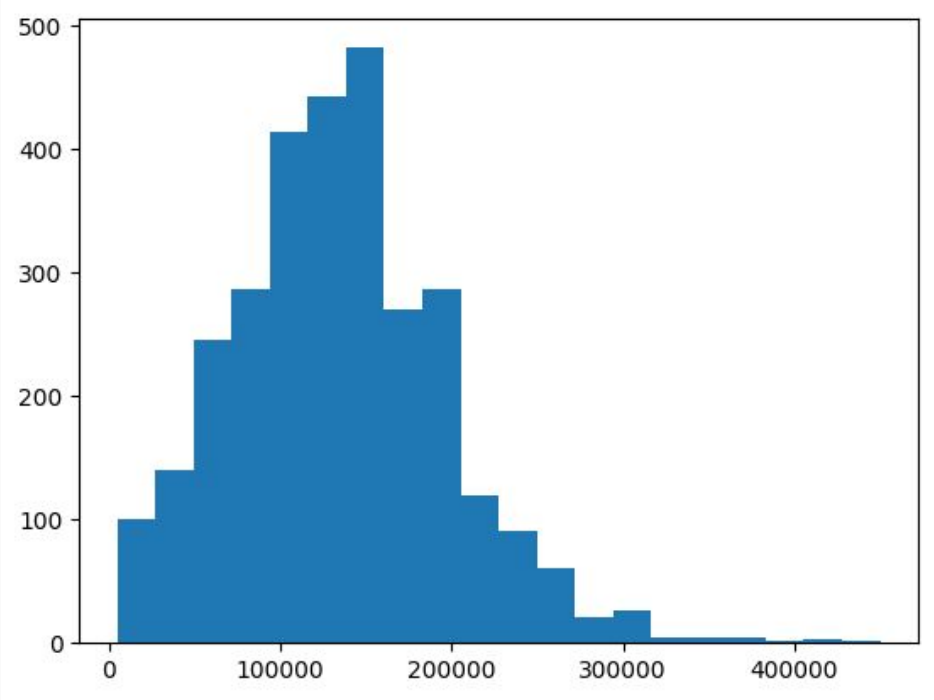
- To make sure they’re not underpaying talent that could leave for a better paying job.
- To make sure they’re not overpaying talent according to their experience level.

Employees:

- Make sure you’re not being paid less than average according to your credentials.
- When applying for jobs have a better idea of the salary range you could demand.

Salary in USD

- The average salary appears to be around $150,000
- Salaries above $400,000 could be errors in the data collection. They will have to be checked individually.



Highest Salaries

	experience_level	employment_type	job_title	salary_in_usd	employee_residence	remote_ratio	company_location	company_size
2773	MI	FT	Research Scientist	450000	US	0	US	M
1262	MI	FT	Data Analyst	430967	GB	0	GB	M
2998	MI	FT	Applied Machine Learning Scientist	423000	US	50	US	L
2926	EX	CT	Principal Data Scientist	416000	US	100	US	S
3001	SE	FT	Data Scientist	412000	US	100	US	L

Considering their experience level, and that their salaries are well above the highest Senior and Executive salaries, we will assume the top 3 salaries are errors in data collection

Relevant Features

Experience Level

Remote Ratio

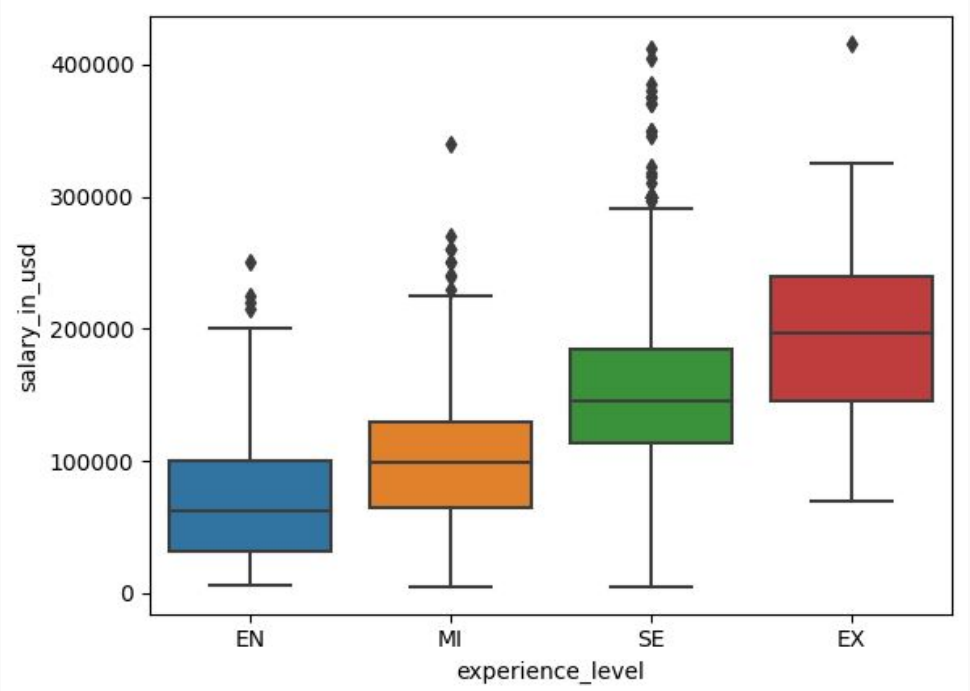
Company Size

Employee Location

Company Location

Experience Level

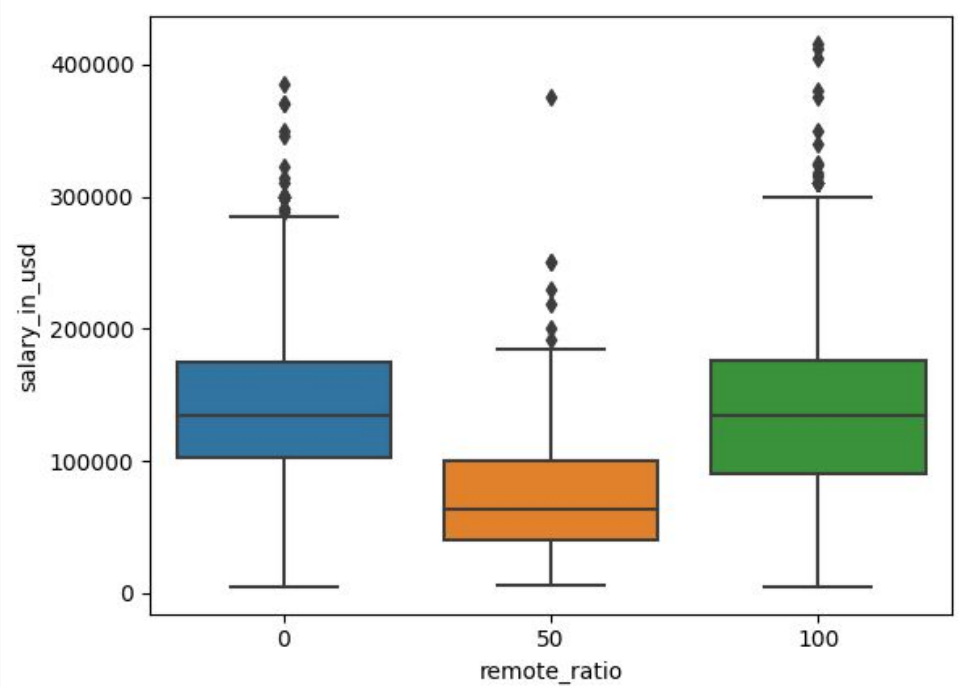
The high outliers are concerning, but the distributions seem to be intuitively correct.



Remote Ratio

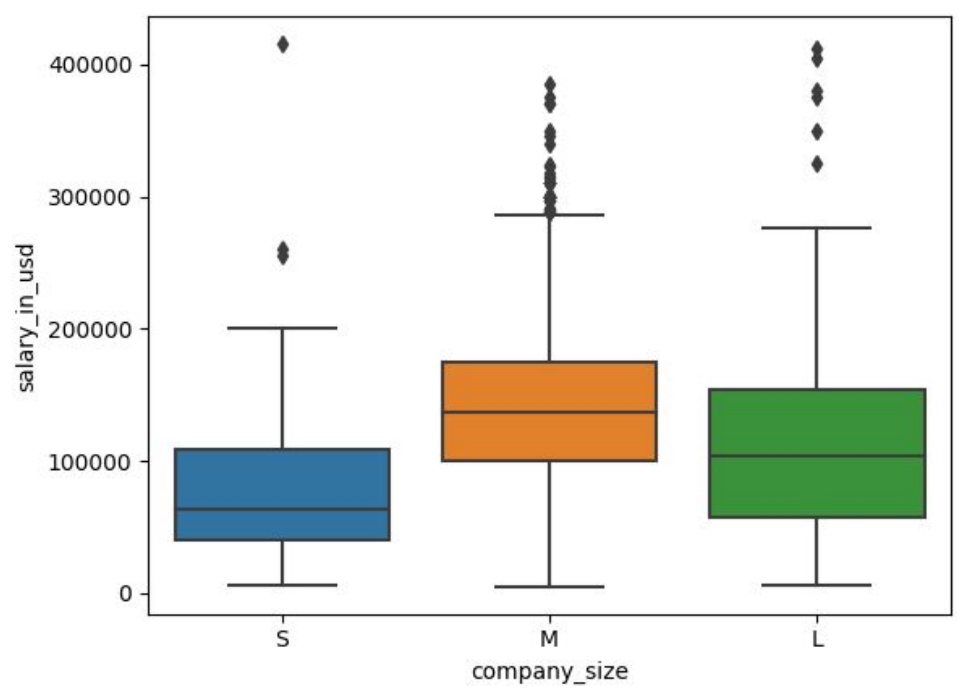
Salaries for remote and in person jobs seem to be very close.

Hybrid workers on the other hand seem to have significantly lower salaries



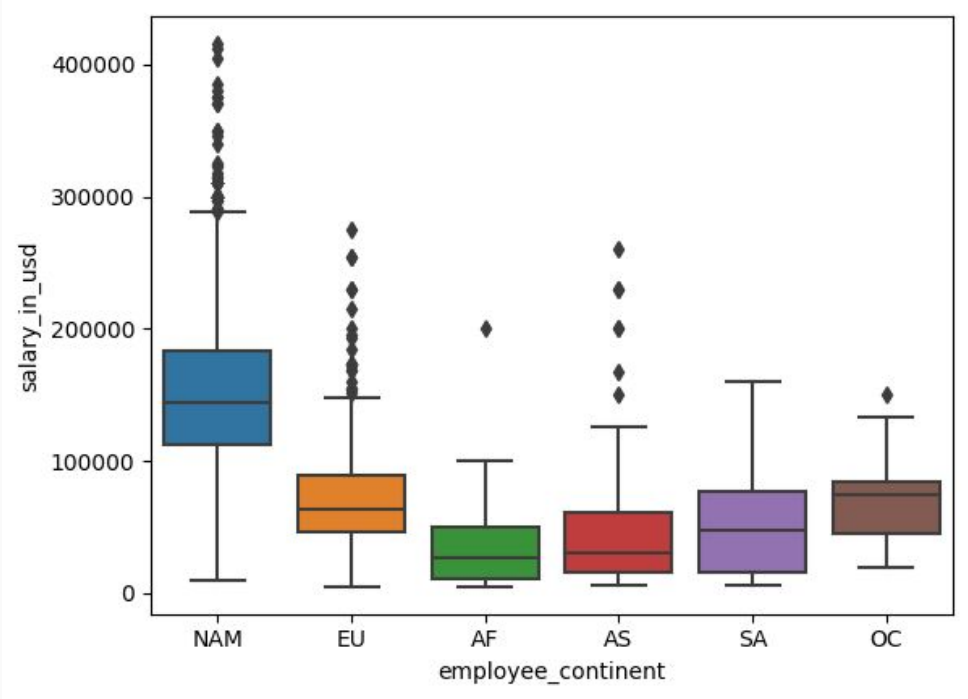
Company Size

On average medium size companies pay the highest salary.



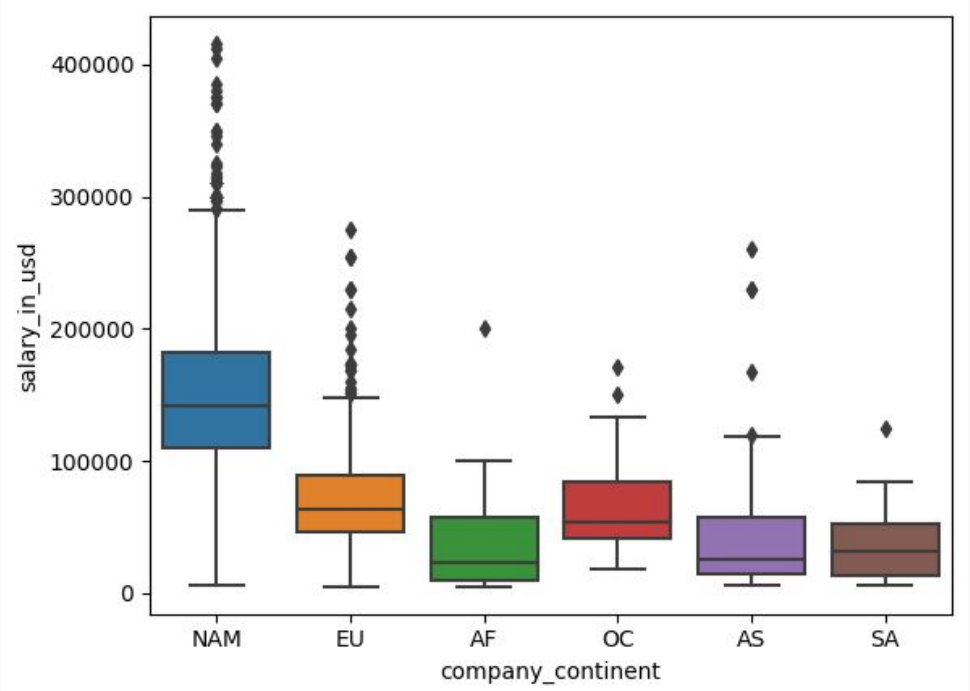
Employee Location

Employees in North America charge the highest salaries, followed by Europe and Oceania.

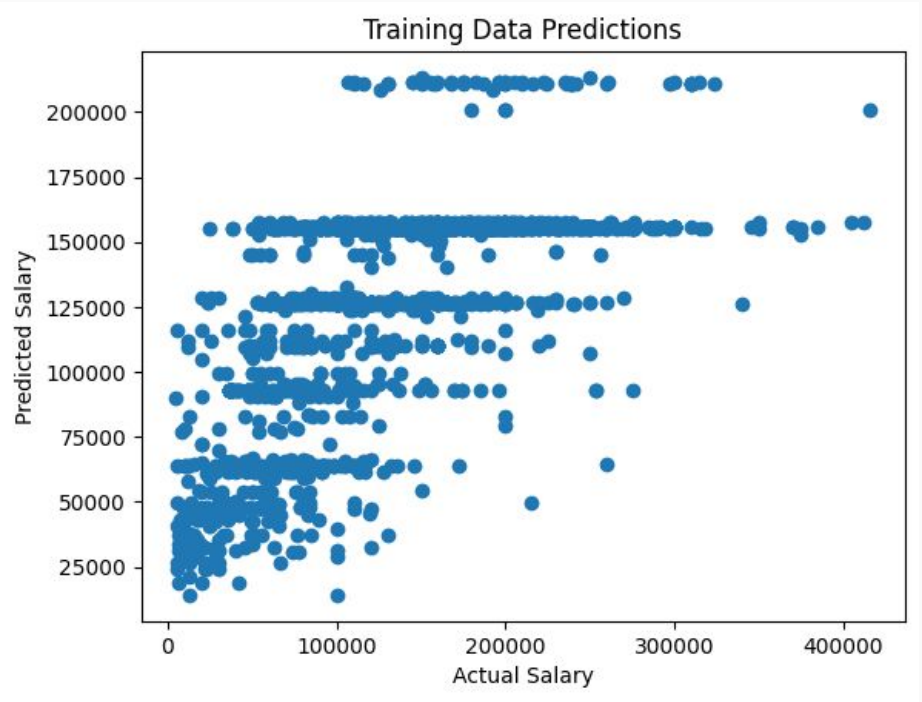


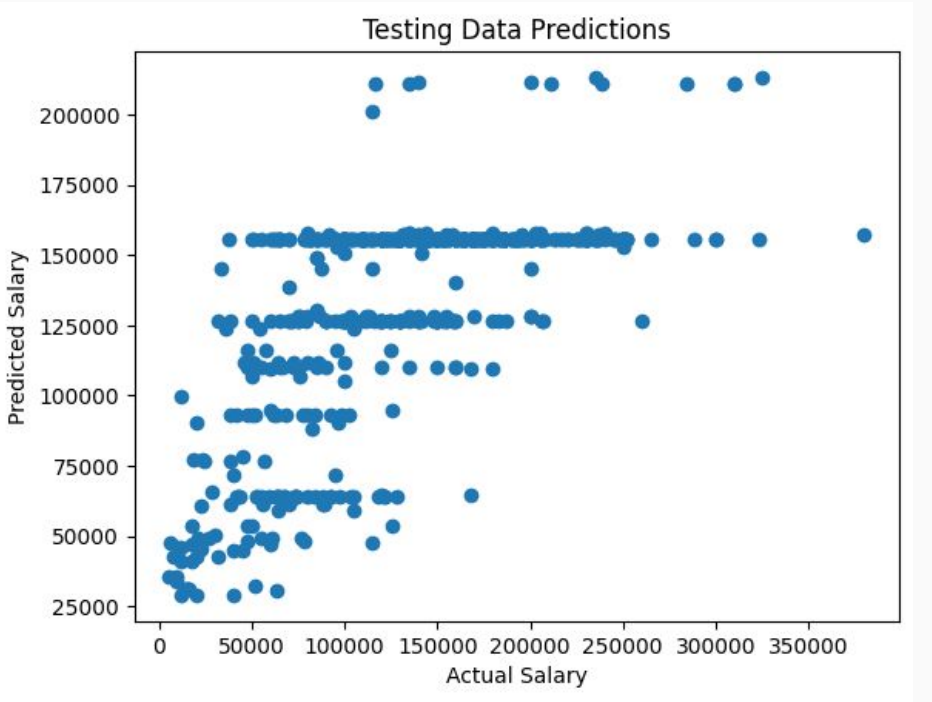
Company Location

Likewise, companies in North America pay the highest salaries, followed by Europe and Oceania.



Linear Regression Model



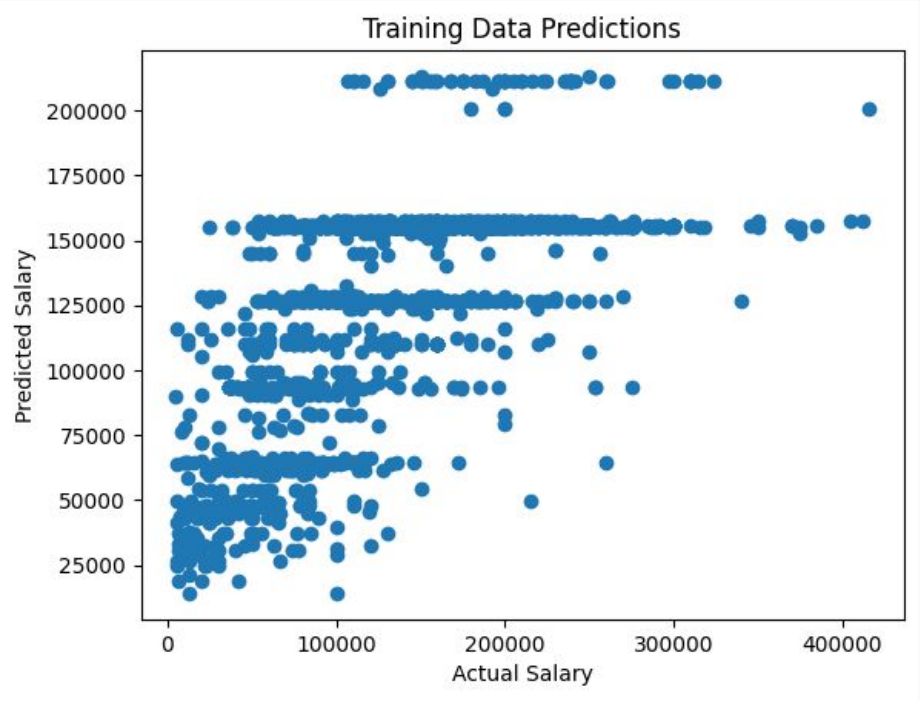


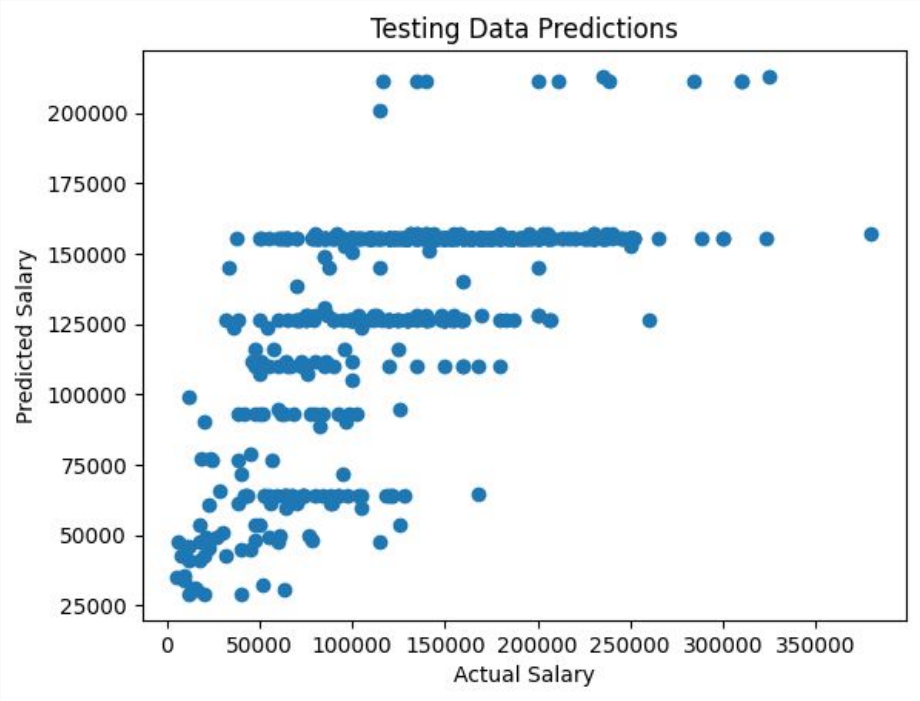
Linear Regression Model

Metrics:

- Training Data R2 Score: 0.370
- Testing Data R2 Score: 0.406
- Mean Absolute Error: $37769.86

Ridge Regression Model



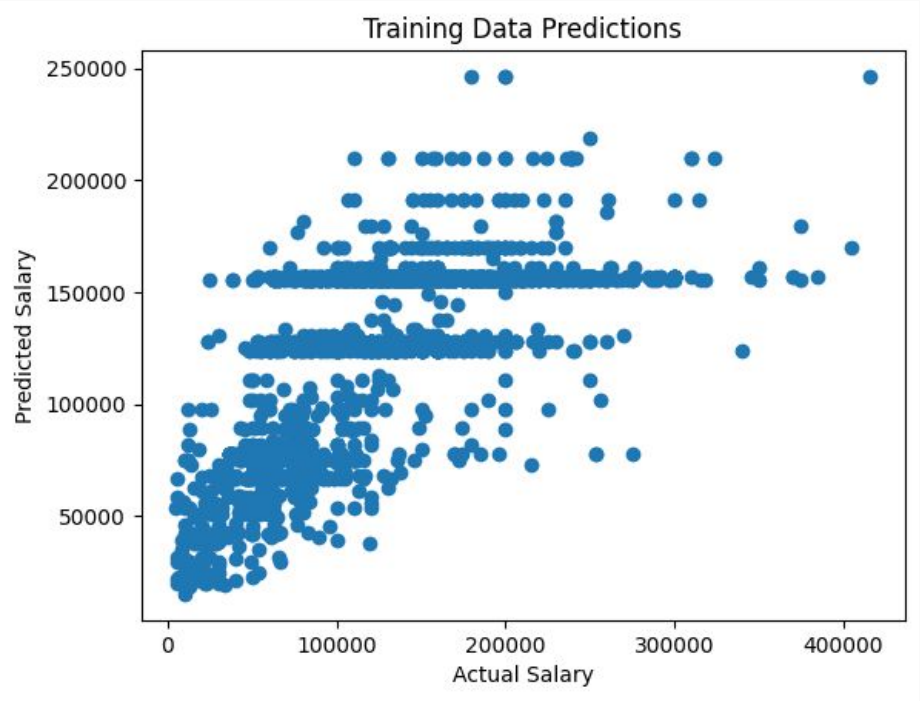


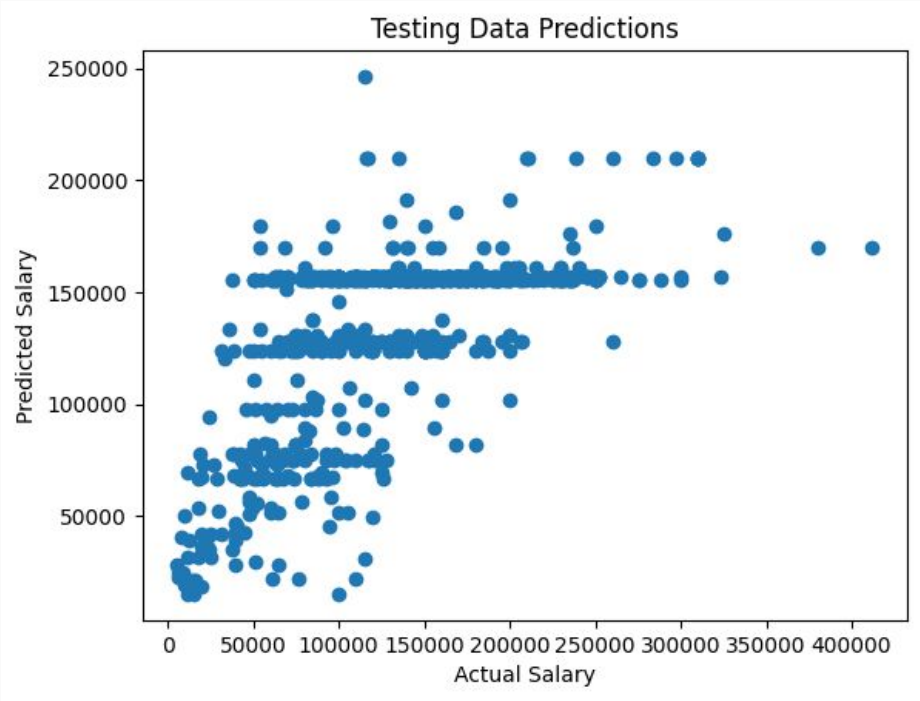
Ridge Regression Model

Metrics:

- Training Data R2 Score: 0.370
- Testing Data R2 Score: 0.406
- Mean Absolute Error: $37769.86

Random Forest Regressor





Random Forest Regressor

Metrics:

- Training Data R2 Score: 0.408
- Testing Data R2 Score: 0.398
- Mean Absolute Error: $37799.36

Model Conclusions

They all produced very similar results with an R-Score of around 0.40, which is not a very accurate model. Our predictions seem to be off by around $37,000 on average.

We can see many of the predictions are boxed in clusters as opposed to the actual salaries. The outliers seem to be affecting the model negatively, as our predictions cap at around $225,000, while the actual salaries go all the way up to $400,000

Future Recommendations

To get more accurate results, we might try getting more entries on our data and make sure the information is accurate, since these are all self-reported, and there could be mistakes when reporting the salary in USD.